Inductive Aspects of Confirmation, Information, and Content

Kuipers, Theo A.F.

Published in:
The Philosophy of Jaakko Hintikka", The Library of Living Philosophers

Publication date:
2006

26

Theo A. F. Kuipers

INDUCTIVE ASPECTS OF CONFIRMATION, INFORMATION, AND CONTENT

INTRODUCTION

This contribution consists of three parts. In section 1, which is a kind of summary of part I of *From Instrumentalism to Constructive Realism*,[1] I will characterize Hintikka's quantitative theory of confirmation as one of the four main theories. Moreover, I will disentangle the *structural* and genuine *inductive* aspects in these theories of confirmation. In section 2 I will develop Hintikka's ideas about two major types of information called *information* and *content*.[2] In 1997 Hintikka wrote, "In hindsight, I regret that I did not develop further the ideas presented in Hintikka (1968)."[3] I will point out the close relation between *transmitted* information and content on the one hand and *confirmation* on the other. Additionally, I will characterize Hintikka's aims and claims regarding the choice between hypotheses when dealing with *explanation* and *generalization* in relation to (transmitted) information and content. In section 3 I will first disentangle the structural and genuine inductive aspects of prior and posterior probabilities and of transmitted information and content, then I will discuss Hintikka's answers to the question of what to maximize in the service of explanation and generalization, viz., maximizing information by maximizing likelihood and maximizing (transmitted) content, respectively. I will suggest alternative answers, viz., in terms of structural and inductive aspects, respectively. I will conclude with some remarks about the choice of a probability function, a problem that tends to remain hidden when dealing with the choice between hypotheses on the basis of a fixed probability function underlying a theory of confirmation.

1. Ordering the Landscape of Confirmation

The aim of this section is to give a coherent survey of qualitative and quantitative notions of confirmation, partly by synthesizing the work of others, in a standard or nonstandard way, and partly by showing the distance between this synthesis and the work of others. I will start with qualitative, more specifically, deductive confirmation, of which the standard form is supplemented with two comparative principles. Keeping deductive confirmation as extreme partial explication in mind, I then turn to quantitative, more specifically, probabilistic confirmation, and introduce the crucial distinction between inductive and noninductive, or structural confirmation. This will lead to a survey of the four main theories of confirmation for universal hypotheses, viz., those of Popper, Carnap, Bayes, and Hintikka. Finally, the section deals with the question of a general (quantitative) degree of confirmation and its decomposition into degrees of structural and inductive confirmation, leading to four kinds of inductive confirmation.

1.1. *Types of Confirmation*

1.1.1. *Deductive Confirmation*

Contrary to many critics, and partly in line with Gemes,[4] I believe that the notion of *deductive (d-)confirmation* makes perfectly good sense as partial explication, provided the classificatory definition is supplemented with some comparative principles. More specifically, "(contingent) evidence E d-confirms (consistent) hypothesis H" is defined by the clause: H (logically) entails E, and further obeys the following principles:

> *Comparative principles*:
>
> P1: if H entails E and E entails E* (and not vice versa) then E d-confirms H more than E*.
>
> P2: if H and H* both entail E then E d-confirms H and H* equally.

To be sure, this definition-with-comparative-supplement only makes sense as a *partial* explication of the intuitive notion of confirmation; it leaves room for nondeductive, in particular, probabilistic extensions, as we will see below. However, let us first look more closely at the comparative principles. They are very reasonable in light of the fact that the deductive definition can be conceived as a (deductive) *success* definition of confirmation: if H entails E, E clearly is a success of H, if not a predictive success, then at least a kind of explanatory success. From this perspective, P1 says that a stronger (deductive) success confirms a hypothesis more than a weaker one, and P2 says that two hypotheses should be equally praised for the same success. In particular P2 runs against standard conceptions. However, in chapter 2 of *From Instrumentalism to Constructive Realism*, I deal extensively with the possible objections and show, moreover, that the present analysis can handle the confirmation paradoxes discovered by Hempel and Goodman.

Deductive confirmation can also be supplemented with "conditional deductive confirmation": E d-confirms H, assuming a condition C, that is, H & C entails E. This type of conditionalization, also applicable for non-deductive confirmation, will not be treated further in this survey.

1.1.2. *Probabilistic Confirmation*

Probabilistic confirmation presupposes, by definition, a probability function, indicated by p, that is, a real-valued function obeying the standard axioms of probability, which may nevertheless be of one kind or another (see below). But first I will briefly deal with the general question of a probabilistic criterion of confirmation. The *standard* (or *forward*) *criterion* for probabilistic confirmation is that the *posterior* probability $p(H \mid E)$ exceeds the *prior* probability $p(H)$ (relative to the background knowledge), that is, $p(H \mid E) > p(H)$. However, this criterion is rather inadequate for 'p-zero' hypotheses. For example, if $p(H) = 0$ and E d-confirms H, this confirmation cannot be seen as an extreme case of probabilistic confirmation, since $p(H \mid E) = p(H) = 0$. However, for p-nonzero hypotheses and assuming $0 < p(E) < 1$, the standard criterion is equivalent to the *backward* or *success* criterion, according to which the so-called *likelihood* $p(E \mid H)$ exceeds the initial probability $p(E)$ of E: $p(E \mid H) > p(E)$. Now it is easy to verify that any probability function respects d-confirmation according to this criterion, since $p(E \mid H) = 1$ when H entails E, and hence exceeds $p(E)$, even if $p(H) = 0$. More generally, the success criterion can apply in all p-zero cases in which $p(E \mid H)$ can nevertheless be meaningfully interpreted.

To be sure, as Maher stresses, the success criterion does not work properly for 'p-zero evidence',[5] e.g., in the case of verification of a real-valued interval hypothesis by a specific value within that interval. However, although this is less problematic,[6] it seems reasonable to accept the standard criterion for p-zero evidence. Note that this leads to the confirmation verdict in the indicated case of verification. From now on "confirmation" will mean forward or backward confirmation when p(H)   p(E), backward confirmation when p(H) = 0 and p(E)   0 and forward confirmation when p(E) = 0 and p(H)   0; and it is left undefined when p(H) = 0 = p(E).

1.1.2.1. *Structural Confirmation*

I now turn to a discussion of the kinds of probability functions and corresponding kinds of probabilistic confirmation. I start with non-inductive or *structural confirmation*, which has an "objective" and a "logical" version.

Consider first an objective example dealing with a fair die. Let E indicate the even (elementary) outcomes 2, 4, 6, and H the "high" outcomes 4, 5, 6. Then (the evidence of) an even outcome confirms the hypothesis of a high outcome according to both criteria, since $p(E \mid H) = p(H \mid E) = 2/3 > 1/2 = p(H) = p(E)$.

I define structural confirmation as confirmation based on a probability function assigning equal and constant probabilities to the elementary outcomes. Such a probability function may either represent an *objective* probability process, such as a fair die, or it may concern the so-called *logical* probability or logical measure function, indicated by m (corresponding to Carnap's c^{+}-function).[7] Kemeny's m-function assigns probabilities on the basis of ([the limit, if it exists, of] the ratio of) the number of structures making a statement true, that is, the number of models of the statement.[8] These logical probabilities may or may not correspond to the objective probabilities of an underlying process, as is the case of a fair die. Hence, for structural confirmation, we may restrict the attention to generalizations of Kemeny's m-function.

Structural confirmation is a straightforward generalization of d-confirmation. For suppose that H entails E. Then

$$m(E \mid H) = (\lim)\ |Mod(E \ \&\ H)| \,/\, |Mod(H)| = \\ 1 > (\lim)\ |Mod(E)| \,/\, |Mod(Tautology)| = m(E),$$

where e.g., '|Mod(H)|' indicates the number of models of H. Hence, structural confirmation might also be called "extended" or "generalized d-confirmation." Moreover, structural confirmation is a probabilistic explication of Salmon's idea of confirmation by "partial entailment," according to which an even outcome of a throw with a fair die typically is partially implied by a high outcome.[9] For this reason we might call nondeductive cases of structural confirmation also cases of "partial d-confirmation."

It is important to note that the m-function leads in many cases to 'm-zero' hypotheses.[10] For instance, every universal generalization "for all *x* F*x*" gets zero m-value for an infinite universe. As we may conclude from the general exposition, such hypotheses may well be structurally confirmed by some evidence, by definition, according to the success criterion, but not according to the standard criterion. For example, a black raven structurally confirms "all ravens are black" according to the success criterion, even if the universe is supposed to be infinite. Typical for the m-function is that it lacks the property, which I will present as characteristic for inductive probability functions.

1.1.2.2. *Inductive Confirmation*

Inductive confirmation is (*pace* Popper and Miller's 1983 paper[11]) explicated in terms of confirmation based on an inductive probability function, i.e., a probability function p having the general feature of "positive relevance," "inductive confirmation" or, as I like to call it, *instantial confirmation*: $p(Fb \mid E \,\&\, Fa) > p(Fb \mid E)$, where 'a' and 'b' represent distinct individuals, 'F' an arbitrary monadic property and 'E' any kind of contingent or tautological evidence. Note that this definition is easy to generalize to *n*-tuples and *n*-ary properties, but I will restrict the attention to monadic ones. Since the m-function satisfies the condition $m(Fb \mid E \,\&\, Fa) = m(Fb \mid E)$, we get for any inductive probability function p:

$$p(Fa \,\&\, Fb \mid E) = p(Fa \mid E) \,.\, p(Fb \mid E \,\&\, Fa) > m(Fa \,\&\, Fb \mid E)$$

as long as we may also assume that

$$p(Fa \mid E) = p(Fb \mid E) = m(Fa \mid E) = m(Fb \mid E).$$

Inductive (probability) functions can be obtained in two ways, which may also be combined. They can be based on:

- "inductive priors." i.e., positive prior values p(H) for m-zero hypotheses

and/or on

- "inductive likelihoods," i.e., likelihood functions $p(E \mid H)$ having the property of instantial confirmation

Note first that forward confirmation of m-zero hypotheses requires inductive priors, whereas backward confirmation of such hypotheses is always possible, assuming that $p(E \mid H)$ can be interpreted. Second, although we now have a definition of inductive probability functions, we do not yet have a general definition of inductive confirmation. In the next subsection I will give such a general definition in terms of degrees of confirmation, but the basic idea is, of course, that the confirmation is (at least partially) due to instantial confirmation.

With reference to the two origins of (the defining property of) inductive probability functions I now can characterize the four main theories of confirmation in philosophy of science:

	inductive priors	inductive likelihoods
Popper	no	no
Carnap	no	yes
Bayes	yes	no
Hintikka	yes	yes

TABLE 1: The Four Main Theories of Confirmation in Philosophy of Science.

Popper rejected both kinds of inductive confirmation, roughly, for three reasons: two problematic ones and a defensible one. The first problematic reason to note (although not convincingly presented) is that p(H) could not be positive.[12] The best positive argument against zero prior probabilities is perhaps that zero priors amount to dogmatic scepticism with respect to the relevant hypotheses.[13] The second problematic argument against inductive confirmation is that any probability function has the property 'p(E   p(E H)'.[14] Although the claimed property is undisputed, the argument that a proper inductive probability function should have the reverse property, since 'E H' is the "inductive conjunct" in the equivalence 'H (E H) & (E H)', is not convincing. The indicated reverse property may well be conceived as an unlucky first attempt to explicate the core of (probabilistic) inductive intuitions, which should be replaced by the property of instantial confirmation. The defensible reason is that the latter property merely reflects a subjective attitude and, usually, not an objective feature of the underlying probability process, if there is such a process at all.

Carnap, following Laplace, favored inductive likelihoods, although he did not reject inductive priors. The so-called Bayesian approach in philosophy of science reflects inductive priors.[15] Finally, Hintikka introduced "double inductive" probability functions, by combining the Carnapian and the Bayesian approach. This leads to "double inductive confirmation," which might well be called "Hintikka-confirmation."[16]

Hintikka (in his "Comment on Theo Kuipers") likes to stress that his approach is not only deviating technically from Carnap and Bayes, but also philosophically, and I agree with the main points. First, in contrast to Carnap, Hintikka does not believe that a convincing logical argument can be given for the choice of the relevant parameters, i.e., nonlogical criteria dealing with subjective attitudes, such as caution, or expectations (e.g., about the order in the universe) are unavoidable. Second, in contrast to Carnap and Bayes, Hintikka believes that there can be good reasons to change the parameters, i.e., to make non-Bayesian moves, as it is sometimes called in the literature. Finally, Hintikka's further ambitions are to take different kinds of information into account when dealing with confirmation (see section 2) and to express "assumptions concerning the orderliness of order in one's universe of discourse . . . by explicit premises rather than [by] choices of the value of a parameter."[17]

1.2. *Degrees of Confirmation*

I now turn to the problem of defining degrees of confirmation, in particular a *degree of inductive confirmation*, even such that it entails a general definition of *inductive confirmation*. The present approach does not follow the letter but is in the spirit of Mura.[18] The idea is to specify a measure for the degree of inductive influence by comparing the relevant *p*-expressions with the corresponding (structural) *m*-expressions in an appropriate way. I will proceed in three stages.

1.2.1. *Stage 1: Degrees of General and Structural Confirmation*

In the first stage I propose, instead of the standard difference measure $\mathrm{p(H \mid E)} \quad \mathrm{p(H)}$, the (nonstandard version of the) ratio measure $\mathrm{p(E \mid H) / p(E)}$ as the degree (or rate) of (backward) confirmation in general (that is, according to some p), indicated by $\mathrm{Cp(H, E)}$. It amounts to the degree of structural confirmation for $\mathrm{p = m}$. This ratio has the following properties: for p-non-zero hypotheses, it is equal to the standard ratio measure $\mathrm{p(H \mid E) / p(H)}$, and hence is symmetric ($\mathrm{Cp(H, E) = Cp(E, H)}$), but it leaves room for confirmation (amounting to: $\mathrm{Cp(H, E) > 1}$) of p-zero hypotheses. (For p-zero evidence we might turn to the standard ratio measure.) Moreover, it satisfies the comparative principles of deductive (d-)confirmation P1 and P2. Note first that $\mathrm{Cp(H, E)}$ is equal to $\mathrm{1 / p(E)}$ when H entails E, for $\mathrm{p(E \mid H) = 1}$ in that case. This immediately implies P2: if H and H* both entail E then $\mathrm{Cp(H, E) = Cp(H^*, E)}$. Moreover, if H entails E and E*, and E entails E* (and not vice versa) then $\mathrm{Cp(H, E) > Cp(H, E^*)}$, as soon as we may even assume that $\mathrm{p(E) < p(E^*)}$. This condition amounts to a slightly weakened version of P1. In agreement with P2 we obtain, for example, that an even outcome with a fair die equally d-confirms the hypotheses {6}, {4,6}, and {2,4,6}, with degree of (structural and deductive) confirmation 2. This result expresses the fact that the outcome probabilities are multiplied by 2, raising them from 1/6 to 1/3, from 1/3 to 2/3, and from 1/2 to 1, respectively. Note also that in the paradigm example of structural nondeductive confirmation, that is, an even outcome confirms a high outcome, the corresponding degree is (2/3) / (1/2) = 4/3. But, in agreement with P1, the stronger outcome {4,6} confirms it more, viz., by degree (2/3) / (1/3) = 2. It even verifies the hypothesis.

As suggested, there are a number of other degrees of confirmation. Fitelson evaluates four of them, among which the logarithmic forward version of our backward ratio measure, in the light of seven arguments or conditions of adequacy as they occur in the literature.[19] The ratio measure fails in five cases. Three of them are directly related to the "pure" character of Cp, that is, its satisfaction of P2.[20] In chapter 2 of *From Instrumentalism to Constructive Realism*, I defend P2 extensively. However, I also argue in chapter 3 of that book, that as soon as one uses the probability calculus, it does not matter very much which "confirmation language" one chooses, for that calculus provides the crucial means for updating the plausibility of a hypothesis in the light of evidence. Hence, the only important point which then remains is always to make clear which confirmation language one has chosen.

1.2.2. *Stage 2: The Degree of Inductive Confirmation*

In the second stage I will define, as announced, the degree of inductive influence in this degree of confirmation, or simply the degree of inductive (backward) confirmation (according to p), as the ratio:

$$R_p(H, E) = \frac{C_p(H, E)}{C_m(H, E)} = \frac{p(E \mid H) / p(E)}{m(E \mid H) / m(E)}$$

A nice direct consequence of this definition is that the total degree of confirmation equals the product of the degree of structural confirmation and the degree of inductive confirmation.

In the following table I summarize the kinds of degrees of confirmation that I have distinguished. For later purposes I also add the log ratio version of the ratio measure and the difference measure and, for enabling easy comparison I mainly list the forward versions of the (log) ratio measures.

	A	B	C	D
1	Notion: n(.)	ratio doc	log ratio doc	difference doc
2	degree of confirmation $c_{\text{-}p}(H; E)$	$c_p(H; E) =_{df}$ $p(E \mid H) / p(E) =$ $p(H \mid E) / p(H) =$ $p(H \,\&\, E) / [p(H)p(E)]$	$c_{lp}(H; E) =_{df}$ $\log p(E \mid H) / p(E) =$ $\log p(H \mid E) / p(H) =$ $\log p(H \mid E) \quad \log p(H)$	$c_{dp}(H; E) =_{df}$ $p(H \mid E) \quad p(H)$
3	Relation between prior and posterior probability and degree of confirmation	$p(H \mid E) =$ $p(H) \times c_p(H; E)$	$\log p(H \mid E) =$ $\log p(H) + c_{lp}(H; E)$	$p(H \mid E) =$ $p(H) + c_{dp}(H; E)$
4	degree of *structural* confirmation $C_{\text{-}m}(H; E)$	$c_m(H; E) =_{df}$ $m(H \mid E) / m(H)$	$c_{lm}(H; E) =_{df}$ $\log m(H \mid E) / m(H)$	$c_{dm}(H; E) =_{df}$ $m(H \mid E) \quad m(H)$
5	degree of *inductive* confirmation $R_{\text{-}p}(H,E)$	$Rp(H, E) = df$ $[p(H \mid E) / p(H)] /$ $[m(H \mid E) / m(H)]$	$Rlp(H, E) = df$ $\log [p(H \mid E) / p(H)]$ $\log [m(H \mid E) / m(H)]$	$Rdp(H, E) = df$ $[p(H \mid E) \quad p(H)]$ $[m(H \mid E) \quad m(H)]$

TABLE 2: Survey of Degrees of Confirmation (doc; ‘p’ fixed).

1.2.3. *Stage 3: Four Kinds of Inductive Confirmation*

In the third and final stage I generally define *inductive confirmation*, that is, E inductively confirms H, of course, by the condition: Rp(H, E) > 1. This definition leads to four interesting possibilities for confirmation according to p.

Assume that Cp(H, E) > 1, that is, assume E confirms H according to p. The *first* possibility is *purely structural* confirmation, that is, Rp(H, E) = 1, in which case the confirmation has *no inductive features*. This trivially holds in general for structural confirmation, but it may occasionally apply to cases of confirmation according to some p different from m. The *second* possibility is that of *purely inductive* confirmation, that is, Cm(H, E) = 1, and hence Rp(H, E) = Cp(H, E). This condition typically applies in the case of (purely) instantial confirmation, since, for example, m(Fa | Fb & E) / m(Fa | E) = 1.

The *third* possibility is that of a combination of structural and inductive confirmation: Cm(H, E) and Cp(H, E) both exceed 1, but the second more than the first. This type of *combined* confirmation typically occurs when a Carnapian inductive probability function is assigned, for example, in the case of a die-like object of which it may not be assumed that it is fair. Starting from equal prior probabilities for the six sides, such a function gradually approaches the observed relative frequencies. If among the even outcomes a high outcome has been observed more often than expected on the basis of equal probability then (only) knowing in addition that the next throw has resulted in an even outcome confirms the hypothesis that it is a high outcome in two ways: structurally (as I showed already in 1.2.1) and inductively. Consider the following example:

> Let n be the total number of throws so far, let n_i indicate the number of throws that have resulted in outcome i (1, . . . , 6). Then the Carnapian probability that the next throw results in i is $(n_i + \ /6) / (n + \ )$, for some fixed finite positive value of the parameter . Hence, the probability that the next throw results in an even outcome is $(n_2 + n_4 + n_6 + \ /2) / (n + \ )$, and the probability that it is ‘even-and-high’ is $(n_4 + n_6 + \ /3) / (n + \ )$. The ratio of the latter to the former is the posterior probability of a high next outcome given that it is even and given the previous outcomes. It is now easy to check that in order to get a degree of confirmation larger than the structural degree, which is 4/3 as I have noted before, this posterior probability should be larger than the corresponding logical probability, which is 2/3. This is the case as soon as $2n_2 < n_4 + n_6$, that is, when the average occurrence of ‘4’ and ‘6’ exceeds that of ‘2’.

It is easy to check that the same example when treated by a Hintikka-system (for references, see section 1.1.2.2.) shows essentially the same combined type of structural and inductive confirmation.

Let me finally turn to the *fourth* and perhaps most surprising possibility: confirmation combined with the "opposite" of inductive confirmation, that is, $R_p(H, E) < 1$, to be called *counterinductive* confirmation. Typical examples arise in the case of deductive confirmation. In this case $R_p(H, E)$ reduces to $m(E) / p(E)$, which may well be smaller than 1. A specific example is the following: let E be Fa & Fb and let p be inductive then E d-confirms "for all *x* F*x*" in a counterinductive way. On second thought, the possibility of, in particular, deductive counterinductive confirmation should not be surprising. Inductive probability functions borrow, as it were, the possibility of inductive confirmation by reducing the available "amount" of possible deductive confirmation. To be precise, deductive confirmation by some E is counterinductive as soon as E gets "inductive load" (see section 3.1) from p, that is, $m(E) < p(E)$.

Table 3 summarizes the distinguished kinds of confirmation: general and deductive.

	A: kind	B: condition	C: degree of confirmation =
	General confirmation	$c_p(H, E) > 1$	
1	*Purely inductive* confirmation: e.g., instantial confirmation	$c_p(H, E) > c_m(H, E) = 1$	degree of inductive confirmation
2	*Combined* (inductive and structural) confirmation:	$c_p(H, E) > c_m(H, E) > 1$	degree of structural confirmation degree of inductive confirmation
3	*Purely structural* confirmation:	$c_p(H, E) = c_m(H, E) > 1$	degree of structural confirmation
4	*Counterinductive* confirmation:	$c_m(H, E) > c_p(H, E) > 1$	degree of structural confirmation degree of "inductive" confirmation
	Deductive confirmation	$H \models E$, and hence $p(E\|H) = m(E\|H) = 1$	$c_p(H, E) = 1 / p(E) =$ $[1/m(E)]\ (m(E)/p(E)) =$ $c_m(H, E)\ R_p(H, E)$
5	*Purely inductive* confirmation	impossible, if $m(E) < 1$	
6	*Combined* (inductive and structural) confirmation	$m(E) > p(E)$	
7	*Purely structural* confirmation	$m(E) = p(E)$	
8	*Counterinductive* confirmation	$m(E) < p(E)$	

TABLE 3: Survey of Kinds of Inductive Confirmation in Terms of Ratio Degrees of Confirmation ('p' Fixed, 'm': Logical Measure Function).

2. Kinds of Information and Their Relevance for Explanation and Generalization

Drawing upon the work of Carnap, Bar-Hillel, Popper, Törnebohm, and Adams, Hintikka has introduced in a systematic way two kinds of information, indicated as surprise value and substantive information, and briefly called information and content.[21] When representing them I will point out the close relation between *transmitted* information and *transmitted* content on the one hand and certain *degrees of confirmation* on the other. Moreover, I will characterize Hintikka's aims and claims regarding the choice between hypotheses when dealing with *explanation* and *generalization* in relation to (transmitted) information and content.

2.1. *Kinds of Information*

2.1.1. *Surprise Value Information*

The first kind of information is related to the surprise value or unexpectedness of a statement. As now is usual in computer science, Hintikka focuses on the logarithmic versions of the relevant notions for their nice additive properties. However, as in the case of confirmation, I prefer and start each time with the nonlogarithmic versions for their conceptual simplicity. I begin with the prior (or absolute), the posterior (or conditional) and the transmitted notions of information. The prior information contained in a statement, e.g., a hypothesis H, is supposed to be inversely related to its probability in the following way:

$$\mathrm{i(H)} =_{\mathrm{df}} 1/\ \mathrm{p(H)}$$

$$\mathrm{inf\ (H)} =_{\mathrm{df}} \log \mathrm{i(H)} = \log 1/\ \mathrm{p(H)} = \quad \log \mathrm{p(H)}$$

The posterior version aims to capture how much information this hypothesis, "adds to" or "exceeds" another, e.g., an evidential statement E, assuming E:

$$\mathrm{i(H \mid E)} =_{\mathrm{df}} 1/\ \mathrm{p(H \mid E)} = \mathrm{p(E)\ /\ p(H\ \&\ E)} = \mathrm{i(H\ \&\ E)\ /\ i(E)}$$

$$\mathrm{inf\ (H \mid E)}^{[22]} =_{\mathrm{df}} \quad \log \mathrm{p(H \mid E)} = \log \mathrm{i(H \mid E)} = \mathrm{inf(H\ \&\ E)} \quad \mathrm{inf(E)}$$

Hintikka calls the inf(H | E) the incremental or conditional information. Now I turn to the idea of information transmission: how much information does E "convey concerning the subject matter" of H, or transmit to H, assuming E?

$$\mathrm{trans\text{-}i(E;\ H)} =_{\mathrm{df}} \mathrm{i(H \mid E)\ /\ i(H)} = \mathrm{p(H \mid E)\ /\ p(H)}$$

$$\mathrm{trans\text{-}inf(E;\ H)}^{[23]} =_{\mathrm{df}} \mathrm{inf(H)} \quad \mathrm{inf(H \mid E)} = \log \mathrm{trans\text{-}i(H;\ E)} = \log \mathrm{p(H \mid E)} \quad \log \mathrm{p(H)}$$

According to Hintikka, we may say that trans-inf(E; H) (and hence trans-i(E; H)) “measures the reduction of our uncertainty concerning H which takes place when we come to know, not H, but E.”[24]

It is directly clear that trans-i(E; H) coincides with the forward ratio measure of confirmation Cp (H; E), and hence that it is symmetric, that is, it equals trans-i(H; E) when p(E) and p(H) are nonzero. However, if p(H) is 0 and p(E) nonzero, trans-i(E; H) is undefined (hence, it may be equated with 1), but trans-i(H; E) may well be defined, viz., when p(E | H) is defined. For instance, in case H entails E trans-i(H; E) = 1/p(E), that is, H reduces the uncertainty regarding E from 1/p(E) to 1 (certainty). Trans-*inf*(E; H) coincides of course with the logarithmic version of the ratio measure, viz., Clp (H; E), and similar remarks apply.

2.1.2. *Substantive Information*

The second kind of information is related to the content or substantive information contained in a statement. Now there are no relevant logarithmic versions. Again I define the prior, the posterior (or conditional) and the transmitted notions of substantive information.

The prior content of a statement, for example, a hypothesis H, is made inversely related to its probability by equating it with the probability of its negation:

cont(H) = df 1 – p(H) = p(H)

It is called the substantive information or the content of H.

For the posterior version, called the “conditional content” by Hintikka, aiming to capture how much information this hypothesis, “adds to” or “exceeds” evidential statement E, assuming E, we get:

cont(H | E)[25] = 1 – p(H | E) = 1 – p(H & E) / p(E)

Now I turn again to the idea of information transmission: how much information does E convey concerning H, or transmit to H, assuming E:

trans-cont(E; H)[26] = df cont(H) – cont(H | E) = p(H | E) – p(H)

According to Hintikka, we may say that trans-cont(E;H) indicates the “*change in the information-carrying status of H which takes place when one comes to know E*.”[27]

It is trivial to see that trans-cont(E; H) coincides with the (forward) difference measure of confirmation. In contrast to the ratio measure (as far as straightforwardly defined), the difference measure is not symmetric, i.e., p(H | E) – p(H) is in general not equal to p(E | H) – p(E). Moreover, if p(H) is 0 and p(E) nonzero, trans-cont(E; H) is 0, that is, the “information-carrying status" of H does not change by coming to know E. This sounds plausible, and so for the reverse: trans-cont(E; H) may well be substantial, viz., when p(E | H) is defined. For instance, in case H entails E trans-cont(H; E) = 1   p(E) = cont(E), that is, the "information-carrying status" of E changes by coming to know H, by cont(E).

Table 4 summarizes the notions introduced so far. I have included Hintikka's main characterizing terms and also the information and content of a conjunction of two independent hypotheses.

	A	B	C	D
1	Notion: n(.)	Information: i(.)	log-information: inf(.)	content: cont(.)
2	Prior value: n(H)	i(H) = df 1/p(H)	inf(H) = df log p(H)	cont(H) = df 1 p(H)
	Hintikka		the *surprise value* or *unexpectedness* of (the truth of) H	the *substantive information* or *content* of H
3	n(H & H′) for independent H and H′, hence, p(H & H′) = p(H) p(H′)	i(H) i(H′)	inf(H) + inf(H′)	cont(H & H′) = cont(H) + cont(H′) cont(H) cont(H′)
4	Posterior value of H given E: n(H | E)	i(H | E) = df 1/p(H | E) = i(H & E) / i(E)	inf(H | E) = df log p(H | E) = inf(H & E) inf(E)	cont (H | E) = df 1 p(H/E) = 1 p(H & E) / p(E)
	Hintikka		*incremental* information (*conditional* information)	*conditional* content
5	Transmitted value (from E to H): trans-n(E; H) corresponds to doc	trans-i(E; H) = df i(H) / i(H | E) = p(H | E) / p(H) = cp(H; E)	trans-inf(E; H) = df inf(H) inf(H | E) = log p(H | E) log p(H) = clp(H; E)	trans-cont(E; H) = df cont(H) cont(H | E) = p(H | E) p(H) = cdp(H; E)
	Hintikka		the information E conveys concerning the subject matter of H	the change in the content of H due to E
	Transmitted value = df prior value / (resp.) posterior value = (log) posterior probability / (resp.) (log) prior probability = df degree of confirmation			

TABLE 4: Survey of Definitions of "Information" and "Content" ('p' Fixed).

In the last row I have also included a compact indication of the plausible relations between the various notions. For the nonlogarithmic version of information we get more specifically:

> transmitted value = df prior value / posterior value =
> posterior probability / prior probability = df
> (forward) ratio degree of confirmation

For the logarithmic version of information we get:

> transmitted value $=_{df}$ prior value   posterior value =
> log posterior probability   log prior probability $=_{df}$
> log ratio degree of confirmation

Finally for the notion of content we get:

> transmitted value $=_{df}$ prior value   posterior value =
> posterior probability   prior probability $=_{df}$
> difference degree of confirmation

Let me illustrate the different notions by the fair die (recall, E/H: even/high outcome), where p = m. For the i-notions we get:

> i(E) = i(H) = 2, i(H | E) = i(E | H) = 3/2, C_m(H; E) =
> trans-i(E; H) = (2/3) / (1/2 = 4/3.

For the inf-notions we simply get the logarithms of these values. For the cont-notions we get: cont(E) = cont(H) = 1/2, cont(H | E) = cont(E | H) = 1/3, C_{dp}(H; E) = trans-cont(E;H) = 2/3   1/2 = 1/6.

2.1.3. *Implication Related Notions*

Strangely enough, Hintikka introduces as a variant of "conditional content" cont(H | E) an alternative definition, called "incremental content," where E is not assumed as "posterior" condition, but only figuring as condition in the implication E   H, viz.,

> $\text{cont}_{\rightarrow}$(H | E)[28] = cont (E   H) = 1 – p(E   H)

Hintikka introduces a separate notation for "cont(E   H)," I am puzzled about why he does not even mention its formal analogue "inf(E   H)," let alone "i(E   H)":

> $i_{\rightarrow}$(H | E) $=_{df}$ i(E   H) = 1 / p(E   H)

and its (logarithmic) inf-version

> $\text{inf}_{\rightarrow}$(H | E) $=_{df}$ inf(E   H) =   log p(E   H)

It is unclear why Hintikka compared, in fact, the pair "inf(H | E)" and "cont(E   H)," even two times (viz., (4) and (5), (6) and (7), respectively). The plausible comparisons seem to be both the pair "inf(E   H)" and "cont(E   H)" and the pair "inf(H | E)" and "cont(H | E)." He only compares the last pair (by (9) and (10)). The only reason for the first mentioned comparison seems to be the "additive" character of both its members (as expressed by (4) and (5)).

For later purposes I introduce Hintikka's definition of the corresponding transmission concept of incremental content.

trans-cont$_{\rightarrow}$(E; H)[29] = df cont(H) cont$_{\rightarrow}$H | E) =
cont$_{\rightarrow}$(H) cont(E H) = 1 p(H v E)

Similarly we can define:

trans-inf$_{\rightarrow}$(E; H) = df inf(H) inf(E H) = log p(E H) / p(H)

and hence

trans-i$_{\rightarrow}$(E; H) = df i(H) / i(E H) = p(E H) / p(H)

Table 5 presents a survey of the implication-related notions.

	A	B	C	D
1	Notion: n(.)	information: i(.)	log-information: inf(.)	content: cont(.)
2	Prior value: n(H)	i(H) = df 1/p(H)	inf(H) = df log p(H)	cont(H) = df 1 p(H)
	Hintikka		the *surprise value* or *unexpectedness* of (the truth of) H	the *substantive information* or *content* of H
3	Value of implication: (E H)	i$_{\rightarrow}$(H; E) = df i(E H) = 1/p(E H)	inf$_{\rightarrow}$(H; E) = df inf(E H) = log p(E H)	cont$_{\rightarrow}$(H; E) = df cont(E H) = 1 p(E H) = cont(H & E) cont(E)
	Hintikka			*incremental* content
4	"Transmitted value by " (from E to H *by E H*): trans-n$_{\rightarrow}$(E; H)	trans-i$_{\rightarrow}$(E; H) = df i(H) / i(E H)	trans-inf$_{\rightarrow}$(E; H) = df inf(H) inf(E H) = log p(E H) / p(H)	trans-cont$_{\rightarrow}$(E; H) = df cont(H) cont$_{\rightarrow}$(H | E) = cont(H) cont(E H) = 1 p(H v E)
	Hintikka			the information E conveys concerning the subject matter of H

TABLE 5: Definitions of "Information" and "Content" Related to the Implication ('p' Fixed).

2.2. *Explanation and Generalization*

From section 7 on, it becomes clear where Hintikka is basically aiming with his distinction between information and content. According to him there is a strong relation with different targets of scientific research. Let me quote him on this matter:

> One of the most important uses that our distinctions have is to show that there are several different ways of looking at the relation of observational data to those hypotheses which are based on them and which perhaps are designed to explain them. In different situations the concept of information can be brought to bear on this relation in entirely different ways. . . . In general, the scientific search for truth is much less of a single-goal enterprise than philosophers usually realize, and suitable distinctions between different senses of information perhaps serve to bring out some of the relevant differences between different goals.
>
> Let us consider some differences between different cases. One of the most important distinctions here is between, on one hand, a case in which we are predominantly interested in a particular body of observations E which we want to explain by means of a suitable hypothesis H, and on the other hand a case in which we have no particular interest in our evidence E but rather want to use it as a stepping-stone to some general theory H which is designed to apply to other matters, too, besides E. We might label these two situations as cases of local and global theorizing, respectively. Often the difference in question can also be characterized as a difference between explanation and generalization, respectively. Perhaps we can even partly characterize the difference between the activities of (local) explanation and (global) theorizing by spelling out (as we shall proceed to do) the differences between the two types of cases.
>
> It is important to realize, however, that in this respect [explanation versus generalizing, TK] the interests of a historian are apt to differ from those of a scientist.[30]

Hintikka argues at length in section 8 that in the cases where we want to explain some evidence E by a hypothesis, we have good reasons, in his own words (replacing symbols): "to choose the explanatory hypothesis H such that it is maximally informative concerning the subject matter with which E deals. Since we know the truth of E already, we are not interested in the substantive information that H carries concerning the truth of E. What we want to do is to find H such that the truth of E is not unexpected, given H."[31] This brings him immediately to a plea for maximizing the transmitted information trans-inf(E; H), which was seen to be equal to log p(E | H) / p(E). For fixed E, this amounts, of course, to support of the so-called *maximum likelihood principle* in this type of case: choose H such that p(E | H) is maximal.

For global theorizing or generalization Hintikka argues extensively in section 9 that we should concentrate on maximizing relevant content notions. As a matter of fact, he has four arguments in favor of the claim that we should choose that hypothesis that maximizes the transmitted content, given E, that is, p(H | E)   p(H). Strangely enough, he only hints upon the only direct argument for this by speaking of "the fourth time" that this expression is to be maximized, whereas he only specifies three, indirect, arguments. However, it is clear that maximizing the transmitted content is an argument in itself, not in the least, because it amounts to maximizing the difference degree of confirmation. But his three indirect arguments are more surprising, or at least two of them are.

These three arguments are all dealing with notions of "expected value gains."[32] First he notes that the "(posterior) expected content gain," plausibly defined by

$$p(H \mid E) \quad cont(H) \quad p(\quad H \mid E) \quad cont(\quad H)$$

is maximal when $p(H \mid E) \quad p(E)$ is maximal. Then he argues that the "expected transmitted content gain," similarly defined by

$$p(H \mid E) \quad \text{trans-cont}(E; H) \quad p(\quad H \mid E) \quad \text{trans-cont}(E; \quad H)$$

is also maximized by the same condition. The latter does not need to surprise us very much because the "expected posterior value gain," viz., $p(H \mid E) \quad cont(H \mid E) \quad p(\quad H \mid E) \quad cont(\quad H \mid E)$, is easily seen to be 0, whereas $cont(H) = \text{trans-cont}(E; H) + cont\ (H \mid E)$.

The third indirect argument is again surprising. It turns out that maximizing $p(H \mid E) \quad p(E)$ also leads to maximizing the "expected transmitted incremental content gain," that is:

$$p(H \mid E) \quad \text{trans-cont}_{\rightarrow}(E; H) \quad p(\quad H \mid E) \quad \text{trans-cont}_{\rightarrow}(E; \quad H)$$

In sum, in the context of generalization, Hintikka has impressive "expectation" arguments in favor of maximizing the transmitted content.

However, I have strong doubts about his considerations. Hintikka's plea for maximization of the transmitted content and hence of the difference degree of confirmation has strange consequences. Consider the case that hypotheses H and H* have equal likelihood in view of E. Suppose, more specifically, that H and H* entail E, and hence that $p(E \mid H) = p(E \mid H^*) = 1$. Then it is easy to check that maximization of $p(H \mid E) \quad p(H) = p(H)(p(E \mid H) / p(E) \quad 1)$ leads to favoring the more probable hypothesis among H and H*. This is a direct illustration of the fact that the difference measure does not satisfy P2 (see section 1.1.1.). Its "impure" character[33] in the form of the Matthew-effect, the more probable hypothesis is rewarded more for the same success, is, surprisingly enough, shared by Popper's favorite measures of corroboration.[34] If any preference is to be expected I would be more inclined to think of the reverse preference in the "context of generalization": if two hypotheses are equally successful with regard to the evidence, then the stronger hypothesis seems more challenging to proceed with. Apart from the fact that Popper would not appreciate the notion of "generalization," this would be very much in Popperian spirit. On the other hand, in the "context of explanation" one might expect a preference for the weaker hypothesis: if one wants to explain something on the basis of the available knowledge one will prefer, among equally successful hypotheses regarding the evidence, the hypothesis that has the highest initial probability, that is, to be precise, the highest updated probability just before the evidence to be explained came available. However, this is not intended to reverse Hintikka's plea into a plea for maximizing likelihood in the case of generalization, for this would, of course, not work in the case of equal likelihoods. But a partial criterion would be possible: when the likelihoods are the same, such as is the case for deductive confirmation, we should favor the weaker hypothesis for explanation purposes and the stronger for generalization purposes.[35] However, even this partial criterion can be questioned because the choice still depends very much on the used probability function. Moreover, it is not easy to see how to proceed in the case of different likelihoods. Simply maximizing likelihood then would discard probability considerations that seem to be relevant at least when the likelihoods do not discriminate.

In the next section I will start to elaborate an alternative view on explanation and generalization, by taking structural and inductive aspects into account. I conclude this section with a totally different critical consideration regarding the notion of "transmitted content." One attractive aspect of the notion of content, whether prior or posterior, is that it has a straightforward qualitative interpretation. If Struct(L) indicates the set of structures of language L and Mod(H) the set of models on which H is true, then the latter's complement Struct(L)   Mod(H) not only represents Mod(¬H) but is also, as Hintikka is well aware, precisely the model theoretic interpretation of Popper's notion of empirical content of H. Hence, cont (H) = p(¬H) may be reconstrued as the p(Mod(¬H)) / p(Struct(L)), where the p(Struct(L)) = 1. Similarly, cont(H | E) may be seen as the probabilistic version of the posterior empirical content of H, viz., Mod(E)   Mod(H). However, I did not succeed in finding a plausible qualitative interpretation of Hintikka's notion of transmitted content. If such an interpretation cannot be given, it raises the question whether "transmitted content" is more than a somewhat arbitrary notion. Note that a similar objection would apply to the notion of "transmitted information" if it should turn out not to be possible to give an interpretation of it in terms of (qualitative) bits of information, of which it is well known that such an interpretation can be given for the underlying notions of prior and posterior information, assuming that 2 is used as the base of the logarithm.

3. Structural and Inductive Aspects of Explanation and Generalization

Whereas Hintikka tries to drive a wedge between explanation and generalization by the distinction between (logarithmic) information and content, in particular the transmitted values, I would like to suggest that the distinction between structural and inductive aspects of these and other notions is at least as important. Hence, first I will disentangle these aspects for some crucial notions in a similar way as I did for confirmation. Next, I will discuss what to maximize in the service of explanation and generalization. Since there do not seem to be in the present context (extra) advantages of the logarithmic version of "surprise value" information, I will only deal with the nonlogarithmic version, from which the logarithmic version can easily be obtained if one so wishes.

3.1. *Structural and Inductive Aspects of Probability and Types of Information*

As we have seen in section 1.2 confirmation has structural and inductive aspects, resulting from a comparison of the values belonging to one's favorite p-function, with the corresponding structural values, that is, the values belonging to the logical measure (m-)function. I will extend this analysis to prior and posterior probabilities and to transmitted information and content. From now on I will call the p-function the "subjective" probability function. However, I hasten to say that "subjective" is not supposed to mean "merely subjective," because such a function may well have been designed in a rather rational way, as Carnap and Hintikka have demonstrated very convincingly.

Let me start by defining the "inductive load" of prior and posterior subjective probability and of the "extra inductive load" of the latter relative to the former. There are of course two versions: a ratio and a difference version. In table 6 I have depicted all of them.

	A	B: ratio loads	C: difference loads
1	Prior inductive load: ni(H)	p(H) / m(H)	p(H) m(H)
2	Posterior inductive load: n_i(H | E)	p(H | E) / m(H | E)	p(H / E) m(H | E)
3	Extra inductive load of posterior probability = Degree of *inductive* confirmation R_p(H; E) / R_{dp} (H; E) = Inductive load in transmitted i / cont	[p(H | E) / m(H | E] / [p(H) / m(H)] = [p(H | E) / p(H)] / [m(H | E) / m(H)]	[p(H | E) m(H | E)] [p(H) m(H)] = [p(H | E) p(H)] [m(H | E) m(H)]

TABLE 6: (Extra) Inductive Loads in Probabilities ('p' Fixed, 'm': Logical Measure Function).

It is easy to check whether the extra inductive load of the posterior probability relative to the prior probability, as indicated in row 3, is equal to the relevant degree of inductive confirmation as defined in section 1.2. However, the latter degrees also equal the ratio or the difference of the transmitted type of information to the corresponding transmitted type of structural information. Hence, it is plausible to define them also as the "inductive load" of the transmitted type of information. Accordingly, for the ratio versions of loads and degrees we get the conceptual relations:

extra inductive (ratio) load (of posterior probability to the prior)
$=_{df}$ posterior inductive load / prior inductive load
= degree of subjective confirmation / degree of structural confirmation
$=_{df}$ degree of inductive confirmation
= transmitted subjective information / transmitted structural information
$=_{df}$ inductive load of transmitted information

And for the difference versions:

extra inductive (difference) load (of posterior probability to the prior)
$=_{df}$ posterior inductive load   prior inductive load
= degree of confirmation   degree of structural confirmation
$=_{df}$ degree of inductive confirmation
= transmitted subjective content   transmitted structural content
$=_{df}$ inductive load of transmitted content

Turning to the question of how to define the (extra) inductive loads of prior and posterior information and content, it is not immediately clear how to proceed. For example, in the difference version it was rather plausible to define the inductive load of the prior subjective probability by p(H)   m(H). In view of the definition of content, 1   p(H), it now seems plausible to define the inductive load in the prior content by [1   p(H)]   [1   m(H)] = m(H)   p(H), and hence equal to "minus the inductive load" in the prior subjective probability. However, one might also argue for equating the inductive (difference) loads of probability and content, or one might define the inductive load as the absolute value | p(H)   m(H) |. Similar considerations and possibilities apply to the ratio versions. However, it is not clear whether we really need all these definitions. Let us see how far we can come by restricting inductive aspects to the probabilities constituting information and content and to the "corresponding" degrees of confirmation, and hence to the transmitted types of information.

3.2. *Explanation and Generalization Reconsidered*

In section 2.2, I discussed Hintikka's arguments in favor of maximizing transmitted information, and hence the (log) ratio degree of confirmation and even the likelihood, for explanation purposes and maximizing the transmitted content, and hence the difference degree of confirmation, for generalization purposes. It turned out that these preferences have strange consequences, in particular in the case of equal likelihoods, notably deductive confirmation. Moreover, Hintikka's perspective seemed rather restricted by taking only one probability function into account. In my opinion at least the subjective and the structural probability function should be taken into account. My claim in this subsection will be that the structural and inductive aspects may be particularly relevant when we want to compare appropriate explanation and generalization strategies. Roughly speaking, I will suggest that inductive features should be minimized in the case of explanation in favor of structural features, and the reverse in the case of generalization.

Our leading question is: What do we want to maximize or minimize in choosing a hypothesis, given the structural m-function and a fixed subjective p-function, when explaining E, and when generalizing from E, respectively? I will certainly not arrive at final answers, but only make a start with answering these questions.[36] The basic candidate values for comparison seem to be the prior and posterior probabilities and the degree of confirmation or, equivalently, the transmitted type of information. In all cases we can compare subjective, structural and inductive values, and in the case of inductive loads and degrees of confirmation we can choose at least between ratio and difference measures. We should keep in mind that maximizing probabilities amounts to minimizing both information and content. In table 7 I have depicted the basic candidate values for maximization or minimization, and indicated Hintikka's preferences for explanation and generalization (indicated in cells 9F and 12F, respectively) and my tentative partial answers (1F and 2F, respectively), which are strongly qualified in the text.

TABLE 7: (next page) Possibilities for Maximizing or Minimizing when Choosing H, for Given p and m (p/m: Subjective/Structural Probability).

	A	B	C	D	E	F
1	prior probability	structural			m(H)	TK: maximize for deductive explanation
2		subjective			p(H)	TK: maximize for inductive generalization
3		inductive load	ratio		p(H) / m(H)	
4			difference		P(H) m(H)	
5	posterior probability	structural			m(H | E)	
6		subjective			p(H | E)	
7		inductive load	ratio		p(H | E) / m(H | E)	
8			difference		p(H | E) m(H | E)	
9	degree of confirmation	structural	ratio	= transmitted structural information	m(H | E) / m(H)	
10			difference	= transmitted structural content	m(H | E) m(H)	
11		subjective	ratio	= transmitted subjective information	p(H | E) / p(H)	JH: maximize when explaining
12			difference	= transmitted subjective content	p(H | E) p(H)	JH: maximize when generalizing
13		inductive	ratio	= extra inductive (ratio) load of posterior probability = inductive load of transmitted information	[p(H | E) / m(H | E] / [p(H) / m(H)] = [p(H | E) / p(H) / [m(H | E) / m(H)]	
14			difference	= extra inductive (diff.) load of posterior probability = inductive load of transmitted content	[p(H | E) m(H | E)] [p(H) m(H)] = [p(H | E) p(H)] [m(H | E) m(H)]	

TABLE 7: Possibilities for Maximizing or Minimizing when Choosing H, for Given p and m (p/m: Subjective/Structural Probability).

Our start will be restricted to the case of equal likelihoods, more specifically, deductive confirmation, and is particularly intended to invite Hintikka to develop his intuitions regarding inductive aspects further. When H and H* both entail E, all relevant likelihoods are simply 1. This result maximizes whatever likelihood will then not work. Moreover, as I have pointed out in section 1.2 (see table 3) deductive confirmation based on p will be counterinductive as soon as p(E) has inductive load in the sense of exceeding m(E). We have also seen in section 2.2 that maximizing the transmitted subjective content, as Hintikka proposed for generalizing, would favor, quite counterintuitively, the weaker hypothesis.[37] But for explanation purposes this might not be so counterintuitive.

So let us first concentrate on explanation. When we merely want to explain (deductively) a certain phenomenon on the basis of the available knowledge, including hypotheses with various probabilities, we would like to be safe. However, instead of saying that this explanation is based on the difference measure of confirmation, I rather prefer to see the preference as essentially based on the following sextuple: p(H), p(E), p(E | H), m(H), m(E), m(E | H) , in terms of which the other notions can be defined. Hence, my tentative answer is that in the context of deductive explanation in the face of deductive confirmation, or more generally in the case of equal likelihoods, the more probable hypothesis should be preferred. This is a robust strategy as far as the comparisons of p- and m-values coincide, or are at least not opposite, which will automatically be the case when one of the hypotheses entails the other. However, the following are among the interesting questions that remain.

> (1) What to do when the comparison of m-values diverges from that of p-values?

I will deal separately with an extreme and a special case, respectively:

> (2) What to do when p(H) = p(H*) = 1 and hence, when both are considered to be established background knowledge, and hence when we will also have m(H) = m(H*) = 1?

> (3) What to do when the m-values of both hypotheses are 0 and the p-values nonzero?

Regarding the first case, I would like to suggest that the hypothesis with the highest prior m-value is the most plausible one to choose. The reason is that the higher subjective prior probability is apparently not due to logico-structural considerations, but due instead to inductive considerations of one kind or another, e.g., order (simplicity, homogeneity) or analogy influences. In the second case, the proper conclusion is that there is apparently more than one valid explanation available. In the third case, and more generally when the prior m-values do not differ, our preferences may well depend on the reason why one subjective prior probability exceeds another. The reason may be due to the fact that one hypothesis is weaker than the other in some logico-structural sense, which cannot be accounted for by the m-function; this may well be the case for m-zero hypotheses. For example, "All A and B are F" and "All C are F" may both get m-value 0, but the first claims in a sense more than the second. In such a case, preferring the one with the higher p-value, due to this aspect, seems plausible for explanation purposes. However, the higher p-value may also be (mainly) due to typical inductive considerations, whereas the relevant hypotheses are logico-structurally comparable. As an example, the one may have more analogical features than the other, accounted for by a higher prior subjective probability. In such a case it will be difficult to choose. However, if the other hypothesis is logico-structurally weaker, but not accounted for by 0 m-values, that one seems preferable again.

Similarly interesting questions remain when I turn to the suggested answer in the case of generalizing: in the context of (inductive) generalization[38] from E by hypotheses that are deductively confirmed by E, the less probable hypothesis should *prima facie* be preferred. However, before I discuss these questions, I have to address the so-called "converse consequence property" of deductive confirmation of H by E, that is, the fact that any stronger hypothesis than H is also deductively confirmed by E, including any one resulting from conjugating H with a totally unrelated additional hypothesis. Of course, we should not prefer such a type of stronger hypothesis. As I have argued in *From Instrumentalism to Constructive Realism* 2.1.2, in such a case the confirmation remains perfectly localizable. Hence, in this case we should only compare hypotheses that are relevant in a sense to be specified, where our limited knowledge of deductive relations should ideally also be taken into account. Assuming that such a definition can be given, the suggestion is that the less probable of the relevant hypotheses should be preferred. Preference here of course means preference for further testing and evaluation, not yet for acceptance, even if that is only for the time being.

Let me now turn to the three remaining questions or cases applied to relevant hypotheses in the context of generalization. Regarding (1), what to do when the p-comparison differs from the m-comparison, I would now like to suggest that the hypothesis with the lowest prior p-value is the most plausible one to choose. The reason is the mirror image of the one in the case of explanation. The lower subjective prior probability is apparently not due to logico-structural considerations but to inductive considerations in favor of other hypotheses; after all, assigning higher probabilities to some hypotheses on the basis of order or analogy considerations has to be paid by other hypotheses. In case (2), the "hypotheses" in question will not be considered as interesting new generalizations because they belong already to the background knowledge. In case (3), when both prior m-values are 0, and more generally when the prior m-values do not differ, our preferences will again depend on the reason why one subjective prior probability exceeds another. It may be due to the fact that the one hypothesis is weaker in some logico-structural sense, which cannot be accounted for by the m-function. In such a case, focussing on the one with the lower p-value, due to this way of being stronger, seems plausible for generalization purposes. However, the lower p-value may also be (mainly) due to typical inductive considerations, working positive for other hypotheses. If another hypothesis is logico-structurally stronger, but is not accounted for by 0 m-values, that one seems now preferable to proceed with.

In sum, as far as choosing between different deductive explanations of new evidence is concerned, my preference goes in the direction of higher structural prior probabilities, with a number of qualifications. On the other hand, as far as choosing between different inductive generalizations is concerned, my preference goes in the direction of lower subjective prior probabilities, with even more qualifications. In both cases the relevant degrees of confirmation and hence the transmitted information do not differ, for the corresponding likelihoods do not differ. I leave the question of how to deal with cases of different likelihoods open.

Concluding Remarks

From sections 2 and 3 it is rather clear that we are far from a final answer to the question how to use probabilities, and measures of confirmation, information and content in the contexts of explanation and generalization. However, I hope to have convinced the reader, and in particular Jaakko Hintikka, whatever the role of different kinds of information, structural and inductive aspects should also play a role. I conclude this paper by enlarging the problem of choices to be made.

It is surprising that Hintikka did not explicitly consider the information-theoretic notion of *entropy*, although he took the corresponding logarithmic notion of information extensively into account. Entropy not only naturally leads to new preference criteria between different hypotheses, but it also suggests preference criteria between different probability functions for the same set of (mutually exclusive and together exhaustive) hypotheses.

Let me start with the first. The *prior entropy* is defined as the prior expected prior (logarithmic) information:

$$p(H)\log p(H) \quad p(\quad H)\log p(\quad H)$$

Similarly, the *posterior entropy* is defined as the posterior expected posterior information:

$$p(H \mid E)\log p(H \mid E) \quad p(\quad H \mid E)\log p(\quad H \mid E)$$

It is plausible to call the difference of the prior entropy minus the posterior entropy, the (amount of) *entropy reduction* due to E. Since entropy measures something like the amount of disorder a hypothesis represents, one plausible option could be to favor the hypothesis that obtains the highest entropy reduction from E. However, I do not have strong feelings in this respect.

So let me turn to the alternative use. Assuming a finite number of mutually exclusive and together exhaustive hypotheses H1, . . . , H*n* ({H, H} forms such a set above), the corresponding prior entropy is of course

$$_{i} \quad p(Hi) \log p(Hi)$$

with similar definitions for the posterior entropy and the entropy reduction. From this perspective the natural question is of course which probability function leads to the highest or lowest prior or posterior entropy and which one to the highest entropy reduction. In fact, the so-called maximum entropy principle, advocated in particular by E. T. Jaynes,[39] is frequently used for selecting the prior distribution with the highest entropy. However, if one is willing to consider non-Bayesian moves, one may of course also consider, in addition, to posteriorily prefer the probability distribution that received the highest entropy reduction.

However this may be, in my opinion, entropy considerations of one kind or another might well turn out to be crucial for finding a satisfactory account of preferences regarding confirmation, information, and content in both the context of explanation and of generalization. That these contexts have to be distinguished carefully I consider to be one of the main challenging and even revolutionary points in Hintikka's contribution to the year 1968.

THEO A. F. KUIPERS

UNIVERSITY OF GRONINGEN
MAY 2001

NOTES

1. Theo A. F. Kuipers, *From Instrumentalism to Constructive Realism*, Synthese Library 287 (Dordrecht: Kluwer, 2000). A provisional version of this section, also entitled "Ordering the Landscape of Confirmation," has appeared as Subsection 7.1.2 (pp. 206–13) of Kuipers, *Structures in Science*, Synthese Library 301 (Dordrecht: Kluwer, 2001). The present version, which appears here with the permission of Kluwer Academic Publishers, underwent some minor revisions and tables 2 and 3 have been added. The section prepares the conceptual ground for sections 2 and 3 which were written for this occasion.

2. See Jaakko Hintikka, "The Varieties of Information and Scientific Explanation," in *Logic, Methodology and Philosophy of Science III*, ed. B. van Rootselaar and J. F. Staal (Amsterdam: North-Holland, 1968), pp. 311–31.

3. Jaakko Hintikka, "Comment on Theo Kuipers," in *Knowledge and Inquiry: Essays on Jaakko Hintikka's Epistemology and Philosophy of Science*, ed. M. Sintonen, Poznan Studies, vol. 51 (Amsterdam: Rodopi, 1997), pp. 317–18.

4. Ken Gemes, "Horwich and Hempel on Hypothetico-Deductivism," *Philosophy of Science* 56 (1990): 609–702.

5. Patrick Maher, "Kuipers on Qualitative Confirmation and the Ravens Paradox," in *Confirmation, Empirical Progress, and Truth Approximation: Essays in Debate with Theo Kuipers*, ed. R. Festa, A. Aliseda, J. Peijnenburg. Poznan Studies (Amsterdam: Rodopi, forthcoming).

6. Theo A. F. Kuipers, "Reply to Patrick Maher," in *Confirmation, Empirical Progress, and Truth Approximation*.

7. John Kemeny, "A Logical Measure Function," *Journal of Symbolic Logic* 18, no. 4 (1953): 289–308.

8. Cf. the random-world or labeled method in Adam Grove, Joseph Halpern, and Daphne Koller, "Asymptotic Conditional Probabilities: The Unary Case," *Journal of Symbolic Logic* 61, no. 1 (1996): 250–75.

9. Wesley Salmon, "Partial Entailment as a Basis for Inductive Logic," in *Essays in Honor of Carl G. Hempel*, ed. N. Rescher (Dordrecht: Reidel, 1969), pp. 47–82.

10. Cf. Kevin Compton, "0-1 Laws in Logic and Combinatorics," in *Proceedings 1987 NATO Adv. Study Inst. on Algorithms and Order*, ed. I. Rival (Dordrecht: Reidel, 1988), pp. 353–83.

11. Karl R. Popper and David Miller, "A Proof of the Impossibility of Inductive Probability," *Nature* 302 (1983): 687–88.

12. See Karl R. Popper, *Logik der Forschung* (Vienna, 1934); translated as *The Logic of Scientific Discovery* (London: Hutchinson, 1959). See also John Earman, *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory* (Cambridge, Mass.: MIT Press, 1992); C. Howson and P. Urbach, *Scientific Reasoning: The Bayesian Approach* (La Salle, Ill.: Open Court, 1989); Theo Kuipers, *Studies in Inductive Probability and Rational Expectation*, Synthese Library 123 (Dordrecht: Reidel, 1978).

13. Cf. Ilkka Niiniluoto, *Critical Scientific Realism* (Oxford: Oxford University Press, 1999), pp. 187–88.

14. Popper and Miller, "A Proof of the Impossibility of Inductive Probability."

15. But Bayesian statistics uses inductive likelihoods as well; see Roberto Festa, *Optimum Inductive Methods: A Study in Inductive Probability Theory, Bayesian Statistics and Verisimilitude* (Dordrecht: Kluwer, 1993).

16. For a survey of "Hintikka-systems" and related systems, such as the systems resulting from the joint work of Hintikka and Ilkka Niiniluoto, see Kuipers, "The Carnap-Hintikka Programme in Inductive Logic," in *Knowledge and Inquiry: Essays on Jaakko Hintikka's Epistemology and Philosophy of Science*, ed. M. Sintonen, Poznan Studies, vol. 51 (Amsterdam: Rodopi, 1997), pp. 87–99, or Kuipers, *From Instrumentalism to Constructive Realism*, section 4.5. For elaborated versions and a study of their properties and relations, see Kuipers, *Studies in Inductive Probability and Rational Expectation*.

17. Hintikka, "Comment on Theo Kuipers," p. 318.

18. Alberto Mura, "When Probabilistic Support Is Inductive," *Philosophy of Science* 57 (1990): 278–89. See also e.g., G. Schlesinger, "Measuring Degrees of Confirmation," *Analysis* 55, no. 3 (1995): 208–12; P. Milne, "*Log[P(h|eb)/P(h|b)]* is the One True Measure of Confirmation," *Philosophy of Science* 63 (1996): 21–26; and Roberto Festa, "Bayesian Confirmation," in *Experience, Reality, and Scientific Explanation*, ed. M. C. Galavotti and A. Pagnini (Dordrecht: Kluwer, 1999), pp. 55–87.

19. Brandon Fitelson, "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science*, supplement to volume 66, no. 3 (1999): S362–S378.

20. The P2-related arguments concern the first and the second argument in Fitelson's table 1, and the second in table 2. Of the other two, the example of "unintuitive" confirmation is rebutted in chapter 3 of *From Instrumentalism to Constructive Realism* with a similar case against the difference measure. The other one is related to the "grue-paradox," for which chapters 2 and 3 claim to present an illuminating analysis in agreement with P2.

21. Hintikka, "The Varieties of Information and Scientific Explanation," in *Logic, Methodology and Philosophy of Science III*. See also Hintikka, "On Proper (Popper?) and Improper Uses of Information in Epistemology," *Theoria* 59 (1993): 158–65. (All further references are to "Varieties," except when otherwise stated.)

22. Here I start to deviate somewhat from Hintikka's notation. Hintikka's "inf_{add}(H | E)," (4), p. 313, and "inf_{cond}(H | E)," (9), p. 315, both correspond with "inf(H | E)," as he notes on p. 315 on the basis of (6) on p. 314.

23. "trans-inf(E; H)" corresponds with Hintikka's notion (11) on p. 316, indicated by him as "transinf(E | H)" on p. 317.

24. P. 316, where I have replaced the relevant symbols.

25. Hintikka's "cont_{cond}(H | E)," (10), p. 315, corresponds with my "cont(H | E)."

26. My "trans-cont(E; H)" corresponds with Hintikka's notion (13) on p. 316, indicated by him as "$\mathrm{transcont}_{cond}$(E | H)" on p. 317; note that I cancelled his sub-

script "cond" because no confusion will arise.

27. P. 316, where I have replaced the relevant symbols.

28. Hintikka's "cont$_{add}$(H | E)," (5), p. 313, corresponds with "cont(E H)," and hence with my cont→(H | E), as he notes by (7) on p. 314.

29. "trans-cont→(E; H)" corresponds with Hintikka's notion (12) on p. 316, indicated by him as "transcont$_{add}$(E | H) on p. 317.

30. Pp. 321, 324 (symbols adapted, TK).

31. P. 321.

32. In section 6 Hintikka writes down the (posterior) expected transmitted values for trans-inf(E; H), trans-cont(E; H) and trans-cont→(E; H) and it would be easy to add those for trans-i(E; H), trans-i→(E; H) trans-inf→(E; H). However, as we will see in the text some of the so-called "expected value gains" are much more interesting.

33. Comparing the "impure" degree $p(H \mid E) – p(H) = p(H)(p(E \mid H) / p(E) – 1)$ with the "pure" degree $p(E \mid H) / p(E)$ makes clear that p(H) can even be considered as a measure of the impurity in the former, at least when the likelihoods are the same.

34. See Kuipers, *From Instrumentalism to Constructive Realism*, appendix 1 of ch. 3.

35. Another option would be to maximize the "expected i-gain," the analogue of expected content gain, which equals $p(H \mid E) / p(H) – p(\neg H \mid E) / p(\neg H) = p(E \mid H) / p(E) – p(E \mid \neg H) / p(E)$, which leads to favoring the one with the lowest likelihood of the negation, also called the degree of *dis*confirmation. This sounds plausible: when the degree of confirmation does not discriminate, we should prefer the hypothesis with the lowest degree of disconfirmation. However, it remains unclear whether this option is more appropriate for explanation rather than for generalization.

36. Moreover, I will neglect the possibility of combining the perspectives of explanation and generalization. This might be done in similar ways as truth value and informativeness have been combined by Isaac Levi for epistemic utility and Ilkka Niiniluoto for truthlikeness (see Niiniluoto, *Critical Scientific Realism*, pp. 168–70). I thank Allard Tamminga for drawing my attention to this type of possibility.

37. Note that when one hypothesis is stronger than the other due to logically entailing the other, comparing m-values will lead to this same conclusion as comparing p-values.

38. Note that using here the phrase "deductive generalization" would be highly misleading, but "inductive generalization" is appropriate, in particular as far as generalizations are concerned that are formulated in the same language as the evidence.

39. See *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, ed. R. D. Rosenkrantz (Dordrecht: Kluwer, 1989).